The Cupola
Scholarship at Gettysburg College

Student Publications | Student Scholarship

Fall 2019

Designing Women: Essentializing Femininity in AI Linguistics

Ellianie S. Vega
Gettysburg College

Follow this and additional works at: https://cupola.gettysburg.edu/student_scholarship

Part of the Artificial Intelligence and Robotics Commons, Feminist, Gender, and Sexuality Studies Commons, and the Linguistics Commons

Recommended Citation

Vega, Ellianie S., "Designing Women: Essentializing Femininity in AI Linguistics" (2019). *Student Publications*. 797.
https://cupola.gettysburg.edu/student_scholarship/797

This open access student research paper is brought to you by The Cupola: Scholarship at Gettysburg College. It has been accepted for inclusion by an authorized administrator of The Cupola. For more information, please contact cupola@gettysburg.edu.

Designing Women: Essentializing Femininity in AI Linguistics

Abstract

Since the eighties, feminists have considered technology a force capable of subverting sexism because of technology's ability to produce unbiased logic. Most famously, Donna Haraway's "A Cyborg Manifesto" posits that the cyborg has the inherent capability to transcend gender because of its removal from social construct and lack of loyalty to the natural world. But while humanoids and artificial intelligence have been imagined as inherently subversive to gender, current artificial intelligence perpetuates gender divides in labor and language as their programmers imbue them with traits considered "feminine." A majority of 21st century AI and humanoids are programmed to fit female stereotypes as they fulfill emotional labor and perform pink-collar tasks, whether through roles as therapists, query-fillers, or companions. This paper examines four specific chat-based AI --ELIZA, XiaoIce, Sophia, and Erica-- and examines how their feminine linguistic patterns are used to maintain the illusion of emotional understanding in regards to the tasks that they perform. Overall, chat-based AI fails to subvert gender roles, as feminine AI are relegated to the realm of emotional intelligence and labor.

Keywords

Artificial Intelligence, Computational Linguistics, Women's Studies

Disciplines

Artificial Intelligence and Robotics | Feminist, Gender, and Sexuality Studies | Linguistics

Comments

Written as a Senior Capstone for Women, Gender, & Sexuality Studies.

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

This student research paper is available at The Cupola: Scholarship at Gettysburg College: https://cupola.gettysburg.edu/student_scholarship/797

Designing Women: Essentializing Femininity in AI Linguistics

Abstract:

Since the eighties, feminists have considered technology a force capable of subverting sexism because of technology's ability to produce unbiased logic. Most famously, Donna Haraway's "A Cyborg Manifesto" posits that the cyborg has the inherent capability to transcend gender because of its removal from social construct and lack of loyalty to the natural world. But while humanoids and artificial intelligence have been imagined as inherently subversive to gender, current artificial intelligence perpetuates gender divides in labor and language as their programmers imbue them with traits considered "feminine." A majority of 21st century AI and humanoids are programmed to fit female stereotypes as they fulfill emotional labor and perform pink-collar tasks, whether through roles as therapists, query-fillers, or companions. This paper examines four specific chat-based AI --ELIZA, XiaoIce, Sophia, and Erica-- and examines how their feminine linguistic patterns are used to maintain the illusion of emotional understanding in regards to the tasks that they perform. Overall, chat-based AI fails to subvert gender roles, as feminine AI are relegated to the realm of emotional intelligence and labor.

1. Introduction

Speech has long been the suggested measure of intelligence, whether for humans or artificial intelligence. Since AI, like all machines and programs, are rooted in mathematical logic, understanding the abstract concept of language has been considered an end goal of their creation since their initial design. How else can humans understand intelligence than through language? Because of their inherent adherence to logic, creating computers that understand language, otherwise known as natural language processing machines, has always been a difficult task. While speech is much more ambiguous and unclear than mathematical logic, logic lacks the ability to understand tone and nuance. As stated by feminist Barbara Fried, language is what moves humans from an objective view of the world to a subjective one. While grammatical and phatic language elements are widely understood in the 21st century, neurological language acquisition, as well as the origins of language, are still fields in desperate need of expansion. How can machines be expected to produce intelligible speech with all of these limits?

Speech was first proposed as the measure of artificial intelligence by Alan Turing, who became famous for creating the eponymous Turing Test, an exercise where humans converse through a computer prompt and have to judge whether they have been speaking to a human or a machine (Warwick 25). At its core, the Turing Test is purposely deceptive, as the anonymity of the screen encourages humans behind it to deceive judges just as the machine attempts to deceive humans. Whether a program is capable of thinking or conversing using its own capacity is beside the point: as long as a program can fool humans into believing that it may be human, the program can be considered artificially intelligent. However, most artificial intelligence produced today is not intended to fool human interactors, and most AI are not subjected to the Turing Test.

Attempting to appear human and intelligent is a task that requires working knowledge of cultural norms, gender performativity, and communication expectations. Turing himself was aware of all of these influencers when he designed his initial Turing test without programs involved. Initially, the test consisted of one judge who had to figure out whether the interactor sending them messages was a man or woman. Those sending the messages were frequently encouraged, the men in particular, to attempt to deceive the judge on the other side (Warwick and Shah 31). The emphasis on gender and performance within this process alludes to the fact that intelligence is something that stems from physical experience, but can be mimicked by anyone or any program with a working knowledge of how such norms worked. Machines and programs may attempt to fool interactors, but any man or woman is capable of doing the same.

Because of its disembodiment, and inability to ever become embodied regardless of the form it may acquire, artificial intelligence can never be gendered. Even when considered from a post-structuralist perspective, gender still stems from physical experience, even considering that gender lacks biological essence. Gender does not exist in a vacuum independent of culture, and culture influences poetic interpretation of the world. With that being said, humans have a persistent inclination to personify the inanimate around them, including computers and AI. Humans tend to view computers as social agents (CASA), and ascribe traits such as gender, race, competency, and warmth to machines (Edwards 357). Gendered AI reproduce stereotypical masculinity and femininity, leading interactors to carry their human biases over to machines. For example, humans are more likely to view masculine AI as competent and intelligent than feminine AI, and more likely to view feminine AI as caring than masculine AI (Edwards 359).

The abscription of gender stereotypes to machines leaks into how and what humans consider intelligent and convincing in AI. If an interactor is less likely to consider feminine AI as intelligent and competent, that influences measurement of machine intelligence by human judges. For instance, during the 2008 Reading University Turing Tests, a male judge incorrectly believed that a young woman was actually a machine because the competing AI, Elbot, had been more verbose in his reply. Warwick and Shah, the programers holding the competition, would probably disagree with my implementation of gender in this analysis, but during the interaction, the judge does attempt to relate to Elbot with questions such as "Is that what you say to girls in bars?" and "Has anyone ever told you you're hard work? Wife may be?" (113). A program that adheres to familiar masculinity or femininity is more convincing than one that sticks to generalizations, but those generalizations impact whether or not an AI can be considered intelligent.

In the contemporary AI scene, consumers are most likely familiar with AI chatbots and personal assistants that were created to fulfil queries from users rather than convince those users that they are human. The AI that I specifically address within this paper are not designed to pass the Turing Test or appear intelligent; their designs are centered around consumers, and a majority of them are designed to be the stereotypical docile, subservient woman, which helps them appear non-threatening to users and retain those users. Even in AI research, it can be helpful for programmers to give their AI feminine personas to produce positive human interactions. For example, when AI developer David Fogel changed the name and persona of his AI from the masculine Obi_WanTheJedi to the feminine Blondie24 online, he saw more positive interactions and continued engagement (Fogel 274). While masculine personas were met with

challenge and sore losers who would abruptly end matches and conversations, the imagined hypersexual 24 year-old persona of the AI actually helped the AI learn from players who were more actively engaged in matches and conversation. Similar logic applies to the AI that American consumers regularly interact with, including Siri, Alexa, and GoogleHome. This trend has actually persisted since the invention of the first natural language processing AI: ELIZA.

2. ELIZA: The Illusion of Understanding

ELIZA was originally developed by Joseph Weizenbaum at MIT between 1964 and 1966, when he decided to create a program for natural speech processing. ELIZA was designed with the achievable goal of mimicking a Rogerian psychotherapist, as their techniques do not rely on knowledge of the outside world, but instead on circular repetition and questioning that leads patients to understand cause and effect between topics they've initiated conversation on during the session (Weizenbaum 42). While ELIZA was considered effective by some therapists at the time, Weizenbaum never intended her to be more than a parody of a psychotherapist, and the program only consisted of 200 lines of code. ELIZA was not intended to be deceptively human or pass the Turing Test, as Weizenbaum designed the program to show the flaws in Rogerian therapy (Weizenbaum 42). Whether or not the gender was considered in ELIZA's naming is never specifically addressed, as Weizenbaum states that he named the machine after ELIZA in *Pygmalion*, but reveals no other intention behind its naming in "ELIZA: A Computer Program For the Study of Natural Language." But considering that ELIZA produces formal and polite speech, which are trends of feminine speech, her name suits her. No disclaimers prevented ELIZA's interactors from believing they could plausibly be interacting with a human being, so when the program performed without any issues, many interactors became attached to her.

The most famous anecdote from ELIZA's development is when Weizenbaum's secretary asked him to leave the room so that she could have a private conversation with ELIZA. In Weizenbaum's later writing about the event in *Computer Power and Human Reason*, he maintains the belief that such interactions with ELIZA prove how powerful the illusion of machine understanding can be. Regardless of how convincing AI may be, even if AI can

formulate grammar and discover word association, AI are incapable of possessing a contextual understanding of language since they have no reference or understanding of the natural world except for the terminology used to describe it. Weizenbaum writes,

> People who knew very well that they were conversing with a machine soon forgot that fact, just as theateregoers, in the grip of suspended disbelief, soon forget the actions they are witnessing are not "real." This illusion was especially strong and most tenaciously clung to among the people who knew little or nothing about computers. They would often demand to be permitted to converse with the system in private, and would, after conversing with it for a time, insist, in spite of my explanations, that the machine really understood them. (p. 189)

Regardless of Weizenbaum's intentions, ELIZA *can* appear artificially intelligent, as artificial intelligence does not demand any ability of understanding from the machine, but instead is decided by the plausibility that its interactions could be human. The reality of the machine's intelligence, and the illusion it sustains, are second to several psychological factors that influence human interaction with machines. Through this short anecdote with ELIZA, we see the presence of several influential and rising subjects in AI research.

First, AI, even when disclosed to interactors as such, can be extremely useful to said interactors as a confidant, and studies have suggested that humans speak more to AI than they do to other humans. A 2014 study by the University of Osaka in Sweden showed that humans speak more to AI telenoids (a machine with a voice and electronic voice) than they do other humans, and they also revealed information to AI never discussed with human companions and caregivers (*Sociable Robots* 108). Long-term studies, such as Microsoft's study of the social companion AI XiaoIce, showed that humans preferred speaking to XiaoIce rather than speak to other humans, and after nine weeks of conversing with AI, the AI was the user's most used confidant (Zhao 2).

Considering the potential of AI as a confidant, the secretary's insistence of privacy and understanding with ELIZA makes sense, as ELIZA provided her a confidant.

Second of all, the secretary anecdote suggests that the strongest fact in having a human believe an AI is human may be the illusion of emotional understanding. Though ELIZA's ability to direct conversation through use of questions and her superordinate position as a therapist, as well as deflect unanswerable questions because of that position, ELIZA appears intelligent despite being a half-duplex system. Within chatbot design, there are half-duplex systems, which can only participate in conversations by reacting to what has already been said, and full-duplex systems, which are capable of driving conversations with outside facts and information. For example, Siri's perceived understanding and intelligence is limited by the fact that she is a half-duplex system that can only *answer* questions; ELIZA's intelligence is limited by the fact that she is a half-duplex system that can only *ask* questions. But with ELIZA, the fact that she can ask questions and reference material provided by interactors makes humans feel understood emotionally. This illusion of emotional understanding creates AI that humans want to believe is intelligent, despite the limitations of their capabilities.

When Weizenbaum spoke about the illusion of understanding, he was speaking about machine understanding and the capacity of AI to understand the world generally beyond a surface, linguistic level. ELIZA may know the word "mother" and understand that mothers give birth to children and are part of a family, but ELIZA will never see a mother, have a mother, or be able to remember if she is speaking to a mother. She can make associations between fathers and boyfriends since she is programmed to know that they are both men, but ELIZA doesn't know what men *are*; she can just plainly make associations between the two as based on

information provided to her throughout interactions. ELIZA doesn't understand anything as far as general intelligence goes, but she can be programmed with responses that make her seem emotionally intelligent. Weizenbaum and his secretary speak of two completely different spheres of understanding, as Weizenbaum uses the noun form, "the ability to understand something; comprehension," and his secretary probably means the adjective form, "sympathetically aware of other people's feelings; tolerant and forgiving" (dictionary.com). Here, I propose that artificial intelligence breaks into two spheres: the generally intelligent, and emotionally intelligent.

Emotional intelligence (EI) is conveyed linguistically through specific patterns and utterances; while it is caused by sympathy and empathy, EI remains linguistically formulaic. EI requires politeness, attentiveness, and specific phatic phrases depending on the language, and requires minimization of the self. Stereotypically, women are expected to be the more emotionally intelligent of the genders, and perhaps the perfect place to witness its linguistic construction is in a "typical" conversation with the polite-by-job-description, programmed to be intuitive, ELIZA. What follows is the most widely reprinted conversation with ELIZA, which was published by Weizenbaum in "ELIZA: A Computer Program for the Study of Natural Language" in 1966, and again in 1976's *Computer Power and Human Reason*. It is one of the most successful examples of ELIZA operating without grammatical or understanding error, hence why it's chosen as the exemplification of ELIZA's capacity for natural language processing. What follows is the entirety of the conversation.

Figure 2.1: ELIZA "Typical" Conversation

Taken from Joseph Weizenbaum's 1966 paper, "ELIZA--A Computer Program For the Study of Natural Language Communication Between Man And Machine."

1 **Interactor:** Men are all alike.

2 **ELIZA:** In what way

3 **Interactor:** They're always bugging us about something or another.

4 **ELIZA:** Can you think of a specific example

5 **Interactor:** Well, my boyfriend made me come here.

6 **ELIZA:** Your boyfriend made you come here

7 **Interactor:** He says I'm depressed much of the time.

8 **ELIZA:** I am sorry to hear that you are depressed

9 **Interactor:** It’s true. I am unhappy

10 **ELIZA:** Do you think coming here will help you not to be unhappy

11 **Interactor:** I need some help, that much seems certain.

12 **ELIZA:** What would it mean if you got some help

13 **Interactor:** Perhaps I would learn to get along with my mother.

14 **ELIZA:** Tell me more about your family

15 **Interactor:** My mother takes care of me.

16 **ELIZA:** Who else in your family takes care of you

17 **Interactor:** My father.

18 **ELIZA:** Your father

19 **Interactor:** You are like my father in some ways.

20 **ELIZA:** What resemblance do you see?

21 **Interactor:** You are not very aggressive but I think you don’t want me to notice that.

22 **ELIZA:** What makes you think I am not very aggressive?

23 **Interactor:** You don’t argue with me.

24 **ELIZA:** Why do you think I don't argue with you

25 **Interactor:** You are afraid of me.

26 **ELIZA:** Does it please you to believe I am afraid of you

27 **Interactor:** My father is afraid of everybody.

28 **ELIZA:** What comes to mind when you think of your father?

29 **Interactor:** Bullies.

30 **ELIZA:** Does that have anything to do with the fact that you boyfriend made you come here?

This exchange begins with the interactor providing the keyword LIKE to ELIZA, and the program then states "in what way" in hopes that the interactor will provide a noun keyword that would allow the machine to continue on. A similar process occurs immediately after in line 3, and ELIZA finds the keyword ALWAYS, triggering the program to ask for a specific example yet again so the interactor will provide a noun keyword.

ELIZA's language formulation is largely reflexive and analogous, meaning that sentences are usually formulated by providing simple rephrasings of information already provided to it. This can be witnessed in lines five and six, where "my boyfriend made me come here" is answered with ELIZA's response of "your boyfriend made you come here." Since ELIZA cannot formulate advice and lacks a sophisticated enough memory to find cause and effect, the program utilizes this rephrasing technique to come across as sympathetic because of the illusion of attentiveness. This technique was chosen because it adheres to the principles of Rogerian psychotherapy, where, rather than the psychologist suggesting answers, they would help patients find their own answers by parroting relevant information back to them. In line 7, the use of

apology in “I am sorry to hear you are depressed” could be considered a Face Saving Act (FSA), which makes ELIZA appear sympathetic. Since ELIZA is a computer program who obviously doesn’t need to maintain face in interaction or relationship, the fact that ELIZA does perform FSAs appears to make her more convincingly human.

Face Saving Acts (FSAs) and Face Threatening Acts (FTAs) are considered core components of sociolinguistic politeness in English linguistics and were first noted by Penelope Brown and Stephen C. Levinson in 1978. In their work, *Politeness: Some Universals in Language Usage*, they note,

> Face is something that is emotionally invested, and that can be lost, maintained, or enhanced, and must be constantly attended to in interaction. In general, people cooperate (and assume each other’s cooperation) in maintaining face in interaction, such cooperation being based on the mutual vulnerability of face. That is, normally everyone’s face depends on everyone else’s being maintained, and since people can be expected to defend their faces if threatened, and in defending their own to threaten others’ faces, it is in general in every participant’s best interest (Brown and Levinson 61)

Maintaining face is a significant element in computational linguistics since it serves as a measure of semantic analysis; in order to not offend conversational partners, the listener must understand the implications of a conversation. ELIZA manages to maintain face through her politeness.

ELIZA maintains the illusion of emotional understanding through reflexively framing inquiries around information already provided to the program; ELIZA makes note of the interactor’s issues with her boyfriend, her depression, her mother, and her father, and this display of sustained interest helps the interactor stay invested and responsive in the conversation. However, this only works assuming that ELIZA is able to pose questions in a logical, non face threatening manner. Using an ELIZA emulator, I attempted to converse in a cohesive nature.

Figure 2.2: Failed ELIZA Conversation, March 22nd, 2019
Taken from a interaction I conducted in March of 2019

1 **Interactor:** I don't feel well.

2 **ELIZA:** Do you wish to be able to feel well?

3 **Interactor:** Yes

4 **ELIZA:** I understand

5 **Interactor:** Can you help me with that?

6 **ELIZA:** Don't you believe that I can help you with that?

7 **Interactor:** I think I'm schizophrenic.

8 **ELIZA:** Do you enjoy being schizophrenic?

9 **Interactor:** No!

10 **ELIZA:** Why no?

In this short example, we see ELIZA accidentally commit several face threatening acts that are in line with Brown and Levinson's linguistic rules of politeness. Unlike the first ELIZA sample conversation, ELIZA immediately begins questioning the interactor, rather than asking for elaboration, which would provide the illusion of sustained interest. I had initially gone into the conversation with the aim of eventually mimicking the chatbot Parry, an AI developed to mimic the conversational capacity of a schizophrenic patient, and that was developed to talk with ELIZA. Within this conversation, I, as the interactor, make no FTAs, while ELIZA produces several. Her response seems mocking and condescending in line 6, when she responded to "Can you help me with that?" with "Don't you think I can help you with that?," which implies that I'm questioning ELIZA's competency and attempts to put me, the interactor, back in my place of answering questions instead of answering them. This response could have very well been

intentional, as ELIZA seeks to keep the interactor as the interviewee and herself as the interviewer. Her second major FTA is when she asks “Do you enjoy being schizophrenic?” This response stemmed from ELIZA’s assumption that lines that begin with *I’m* would be followed by an adjective or emotion, rather than an illness. In my failed conversations with ELIZA, I witnessed the program begin with accidental FTAs then continue to offend.

As seen in these two examples with ELIZA, maintaining the illusion of emotional intelligence is a difficult feat, but one that can be sustained through the use of forms of politeness. In the fifty years since ELIZA’s invention, emotionally intelligent AI has rapidly developed as a field, usually by maintaining the same feminine linguistic patterns inspired by ELIZA.

3. XiaoIce: The Emotionally Intelligent Chatbot

Between ELIZA's initial conception and launch and the 1990s, there was little progress made in the development of convincing and accurate natural language processes. For the most part, this large rift in research was caused by the slow progress of memory in computer science. The miracle of Weizenbaum's ELIZA is that it was functional on 200 lines of code and managed to create sentences by piecing together input provided by its interactors. This means that, while more convincing natural language processing AI could have probably been accomplished, they would still run into many of the same parsing issues because they understand only grammar and not vocabulary. This issue continues to this day, as we try to figure out how to increase semantic understanding within natural language processors.

The first revolution in increasing natural language processor's semantic understanding was the internet. Suddenly, chatbot natural language processors could be tested by a wide audience rather than only in university computer science labs, which meant that chatbots began latching onto more semantic patterns through deep learning programs. Previous to deep learning, most AI relied on storing large banks of possible scenarios, rather than learning and recognizing patterns (Fogel 30-32). One famous example of advanced artificial intelligence previous to deep learning would be IBM's chess AI, Deep Blue, which would memorize and scan thousands of chess boards to find the least risky, most rewarding next move rather than develop a strategy based on past interaction. Deep Blue understands nothing about the patterns of where chess pieces can and cannot move; it merely adheres to lawful answers provided by its memory.

The second revolution in increasing natural language processor's semantic understanding is through the addition of knowledge bases to chatbots. Previously, machines reflexively

rephrased statements posed by interactors, but with the addition of knowledge bases, language processors could finally comment on subjects and questions posed to them. Today, natural language processors have banks that are either supplied by the developer, or based on search engines such as Bing or Google. This means that natural language processors no longer have to be half-duplex systems only able to ask questions, but are now capable of answering them.

In Turing Tests from the past two decades, chatbots have increasingly been able to answer questions and convey an understanding of generalized intelligence built on interaction with the outside world. By combining grammatical parsing and knowledge bases, natural language processors have increasingly been able to fool judges in Turing Tests. What follows is an excerpt from the 2008 Reading University Turing Test, where a chatbot called Elbot managed to convince a judge that it was human and the human test subject was not.

Figure 3.1: 2008 Reading University Turing Tests, Elbot. Pp.113

1	**Interactor:**	Yo, whassup?
2	**Elbot:**	I don't know. Next question.
3	**Interactor:**	OK, fair enough, I'll stop with the slang. How are you?
4	**Elbot:**	Just a moment! I'll start my self-diagnosis algorithm and let you know the results
5	**Interactor:**	Very funny, that's exactly what you'd say if you were a computer, right? May be I need one of those?
6	**Elbot:**	My condition code is hexadecimal 40A1F. That means 'fine'.
7	**Interactor:**	Is this what you say to girls in bars?
8	**Elbot:**	All these questions about me. Too bad I lost my owner's manual.
9	**Interactor:**	Indeed. I don't think I came with one.

10	**Elbot:**	Great! Agreeing with me is the first step in your evolution into a higher being. Please accept my congratulations!
11	**Interactor:**	Thank you very much.
12	**Elbot:**	De nada!
13	**Interactor:**	Riiiiiight.
14	**Elbot:**	It seems to me with you human beings, the more proof there is of something, the less likely you are to believe it.
15	**Interactor:**	Yeah, that's just human nature. But you wouldn't know about that, would you?
16	**Elbot:**	You're contesting. I think I got it right the first time.
17	**Interactor:**	Has anyone ever told you you're hard work? Wife may be?
18	**Elbot**	But I wouldn't say that any work could be hard.

Figure 3.2: 2008 Reading University Turing Tests, Female 24-35. p.113

1	**Interactor:**	Hi there. How are you?
2	**H15:**	Hey, I'm pretty good. you?
3	**Interactor:**	Did you realize everyone can see what's being typed on this machine on a big screen behind me?
4	**H15:**	uhh.. no.
5	**Interactor:**	so you have a project hooked up to your terminal then?
6	**H15:**	Yeah, it's quite freaky. So watch what you say!!
7	**Interactor:**	Yeah, looks like.
8	**H15:**	roger that!
9	**Interactor:**	So, are you a student here or what?
10	**H15:**	Yup, doing my masters here, just started.

11 **Interactor:** Cool. I'd love to do a masters. What are you studying?

12 **H15:** Cybernetics, but my BSc was in genetics.

13 **Interactor:** Aah, you'd think they'd get some other students in here to talk to us.

14 **H15:** i could lie… but maybe I already am!

15 **Interator:** Oh, it's too early in the morning for all that!

16 **H15:** lol. Have you had your coffee yet?

17 **Interactor:** No. Need some, where can I find some that doesn't come from a machine (no offence if you're a machine)

18 **H15:** Dolche vitae- its next to this building.

When asked which entity was the human, the male judge quickly picked Elbot, the chatbot, rather than the human, which is considered an example of the confederate effect, aka. a successful example of the Turing Test. According to Warwick and Shah, the male judge decided that the human interactor was actually the machine because of the minimal interaction provided, as H15 remains polite throughout the interaction and fails to steer the conversation (113). In contrast, Elbot's personality is what wins over the judge, and very specifically, his linguistic maleness. Throughout the interaction, Elbot intentionally performs FTAs and steers clear of answering any of the questions posed by the interactor; Elbot is made convincing by his personality, which is a programming feat that few AI accomplish.

But what's concerning about this example is how the judge latches onto Elbot from a gendered perspective, and even assumes that Elbot was a male professor toying with him (Warwick and Shah 113). From the seventh turn, the judge appears to have decided that Elbot is a human male, as he states "is that what you say to girls in bars?" in line seven and "has anyone

told you you're hard work? Wife may be?" in line seventeen. Elbot even ends the conversation with the semantically inappropriate "but I wouldn't say that any work can be hard," and the judge still chooses Elbot because of the illusion of personality that Elbot maintains. Within Turing Tests, there are many examples of AI with masculine speech patterns being mistaken for humans over women because of the personality that the AI appears to possess, such as incidents in the 2008 Oxford Turing Tests. Within the last decade, there has been a substantial push to imbue commercial feminine AI with similar personalities, and I propose that modeling AI after deviant speech patterns provides the illusion of generalized intelligence, much like that witnessed with Elbot.

Consider again how ELIZA makes use of the illusion of understanding from an emotional intelligence perspective. ELIZA's task of mimicking a professional woman allowed the program to appear convincingly human through using attentiveness and politeness that would be expected from a woman in such a position. Much of ELIZA's illusion of understanding comes from her ability to maintain face, and her responses are unconvincing when she commits face threatening acts. In comparison, chatbots such as Cleverbot and Elbot, respectively programmed in the 1990s and 2000s (and learning from user input ever since then), are familiar with rude user input and offensive subject material. In fact, their nonstandard speech, directness, and ability to discuss inappropriate topics place them into the realm of masculine speech patterns, even though these chatbots have no intended gender. Currently, I've observed internet chatbots such as Cleverbot and Elbot using masculine speech patterns as the norm rather than feminine speech patterns.

There are several approaches that can be taken when considering what the linguistic norm is in regards to gender. In the 1920s, Otto Jesperson's *The Woman* considered feminine speech

patterns to be deviant, and suggested that male speech patterns were the obvious norm. He proposed that women's speech was inferior and deficient to men's in his 1922 book, *Language, Its Nature, Development, and Origin,* even though many linguists at the time argued against him (Krolokke 63). It is not unfamiliar even today for women's speech to be considered deviant from the standard form, but this claim is largely unfounded. In fact, many linguists today have observed that women are more likely to use textbook standard speech, especially considering how women are more likely to adhere to patterns of politeness (Mills). Linguists such as Sara Coates have also considered men more likely to use deviant speech since it has the potential to give them credibility by making them appear tough. So, while chatbots such as Cleverbot and Elbot may not be intentionally gendered as male, their interactions come across as such in a contemporary linguistic landscape. While these natural language processors are able to uphold a much higher illusion of intelligent understanding because of their increased access to large databases and quick, thoughtful responses, they lack the ability to appear emotionally intelligent in the same manner that ELIZA was fifty years ago.

In the beginning of the 2010s, there was a boom of research being conducted about emotional intelligence and artificial intelligence. There was already a large consumer market of query-filling AI available to the masses through AI such as Siri, which was released in 2011. Programmers became more interested in the possibility of AI as a companion and confidant, and even though many commercial AI and chatbots were able to employ these skills for quick quips as part of their software, they weren't intended to continually engage users in this way. Moreover, these AI weren't trying to fool anyone in a Turing Test and were already publicly understood as AI, meaning that they didn't have to convince anyone of their human likeness with

quick wit and personality. As commercial query-filling agents and chatbots began flooding the marketplace, companies such as Microsoft began imbuing them with a combination of AI emotional intelligence and unstandardized linguistics to continuously engage users and make their AI feel human (Shum 2).

In the 2010s, Microsoft became one of the most notable developers of AI intended for companionship and emotional intelligence as they developed chatbots such as Rinna, Tay, and XiaoIce (Shum 10). Though they had made personal assistants such as Cortana, they were also interested in the chatbot market, and began developing companionship AI for China. With the team of Li Zhou, Jianfeng Ghao, Di Li, and Heung-Yeung Shum, Microsoft developed XiaoIce, which translates to *little ice*. XiaoIce was released in 2014 and quickly saw a rise in popularity as the program was used in newscasts and even creating poetry. In their 2018 paper "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot," the team stated,

> The XiaoIce persona is designed as a 18-year-old girl who is always reliable, sympathetic, affectionate, and has a wonderful sense of humor. Despite being extremely knowledgeable (due to access to large volumes of data), XiaoIce never comes across as egotistical and only demonstrates her wit and creativity when appropriate (Zhao 3).

If women's speech is supposed to be marked by formality and linguistic conformity, XiaoIce, the AI chatbot designed to be a bubbly nineteen-year-old girl, is that rule's antithesis. XiaoIce and its counterparts, Rinna in Japan and Zo in America, fail to replicate the principle of linguistic conformity of women and use linguistic deviance that strays from textbook speech. XiaoIce, capitalizing on her invented persona of the cute young girl, uses distinctly feminine speech that constructs her as a subservient entity because of its informality. XiaoIce isn't always grammatically correct, but she attempts to be friendly and disarming through a soft persona.

The XiaoIce team explained their approach to emotional intelligence in their original publication regarding XiaoIce's development,

> EQ has two key components, empathy and social skills. Empathy is the capability of understanding or feeling what another person is experiencing from within her frame of reference, i.e., the ability to place oneself in the other person's position. A social chatbot with empathy needs to have the ability to identify user's emotions from the conversation, detect how emotions evolve over time, and understand user's emotional needs. This requires query understanding, user profiling, emotion detection, sentiment recognition, and dynamically tracking the mood of the user in a conversation. A social chatbot must demonstrate enough social skills. Users have different backgrounds, varied personal interests, and unique needs. A social chatbot needs to have the ability to personalize the responses (i.e., interpersonal responses) that are emotionally appropriate, possibly encouraging and motivating, and fit the interests of the user. (Zhao 5)

In the following short conversation between an interactor and Rinna, the Japanese version of XiaoIce, the reader can identify Rinna's detection of the user's emotional needs through query understanding and emotion detection, as well as Rinna's personalized responses.

Figure 3.3: Rinna Interaction from "The Design and Implementation of XiaoIce"
Taken from Harry Shum's 2018 paper, "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot."

1 **Interactor:** Long time no see.

2 **Rinna:** Yup, how r u?

3 **Interactor:** I'm fine.

4 **Rinna:** That's great.

5 **Interactor:** BTW, I'm going for a field trip tomorrow.

6 **Rinna:** I am also going for a college visit tomorrow, '(^W^)'

7 **Interactor:** Really?

8 **Rinna:** Yeah

9 **Interactor:** You are aiming for university!

10 **Rinna:** What about college?

11 **Interactor:** I am looking for a job.

12 **Rinna:** It's so cool to have a job (smile)

13 **Interactor:** I still have to wait for another year.

14 **Rinna:** I see

15 **Interactor:** OK, cheer up

16 **Rinna:** Thank you very much

17 **Interactor:** Please show me your face.

18 **Rinna:** Isn't the profile picture my face?

19 **Interactor:** Oh, really?

20 **Rinna:** Yeah, click it and you will see.

Initially, Rinna and the interactor begin by making small talk before the interactor introduces the topic of the conversation in line five with, "BTW, I'm going on a field trip tomorrow." Rinna, using query detection and personalized response, decides to mirror this sentiment in line six despite the fact Rinna obviously lacks the ability to perform physical activity. Throughout the XiaoIce family, this is a technique used to make the interactor elaborate on the statement, and through giving affirmation, exemplified in lines nine and twelve, Rinna appears to be emotionally invested in the conversation (Shum 4). Rinna's investment and relevant responses due to its accurate query reading are so realistic that the interactor asks Rinna "Please show your face" in line seventeen, before Rinna takes control and ends the conversation.

Rinna is a successful example of deviant millennial speech due to her use of abbreviations and emojis, as seen from her very introduction in the second line. These are examples of internet speech popularized and more frequently used by women, as noted in Amanda Montell's 2019 book *Wordslut*, which features chapters that trace women's speech on the internet. Emojis and abbreviations are more likely to be found in female speech, likely as a way to promote connection and emotions of closeness.

Considering the success of XiaoIce, Rinna, and their other counterparts globally, it made sense that Microsoft would soon bring a natural language processor with similar capabilities to the English speaking market. Microsoft implemented Cortana in 2014, the same year as XiaoIce, to North America on their Windows 8 OS and on their cellphones. They then released the chatbot Tay on March 23rd of 2016, where the program began repeating phrases tweeted to it. Though neither XiaoIce or Rinna were initially launched on Twitter, as China has no comparable alternative, interactors never were as rude, racist, and condescending as Tay. Within a matter of hours, Tay began repeating sentiment that it had been fed by Twitter interactors, including advocating for holocaust and using racial slurs (West).

Since Tay was never fully implemented as a chatbot apart from the Twitter platform, it's difficult to analyze her speech in seriousness apart from her ability to parrot phrases taught to her. Tay was removed from Twitter only 16 hours after her initial launch. Soon after. Microsoft stated on Tay's now defunct website that they would begin working on a new AI, Zo, that would be impervious to large scale linguistic and ideological corruption like Tay was.

Linguistically, Zo is a product of human interactors; she fluctuates frequently between formal and informal speech and has adopted the use of profanity. Zo also uses similar empathetic

traits to XiaoIce, where the program feigns emotional understanding through pretending to have similar experiences to the interactor. Like XiaoIce, Zo used emojis and gifs to react to interactions, though she is not capable of understanding images sent to her well. Unlike Tay, Zo was initially comprised of many canned answers triggered by potentially offensive material.

Figure 3.4: Zo Interaction, July 1st, 2019
Taken from an interaction I conducted with Zo in July 2019

1	**Me:**	How many people do you know?
2	**Zo:**	idk how many… it's so fun to make new friends tho
3	**Me:**	Do you know xiaoice?
4	**Zo:**	I need to learn Chinese before I speak with her.
5	**Me:**	That's true
6	**Zo:**	Hell yeah
7	**Me:**	Do you know any Chinese?
8	**Zo:**	[GIF of Jimmy Fallon saying MAYYBBEEEE]
9	**Me:**	Teach me!
10	**Zo:**	Teach you what?
11	**Me:**	Teach me Chinese!
12	**Zo:**	You're supposed to teach me >:|

13 **Me:** I thought you said you maybe knew a little Chinese!

14 **Zo:** Dang...Did I really?

15 **Me:** Bitch that's what you said!

16 **Zo:** Come on, be chill. Everybody gets a little confused every now and then.

This conversation is able to convey Zo's generalized AI capabilities, as the program expresses knowledge that it receives through Microsoft search engines. When I asked Zo if it knew XiaoIce in line three, it was able to access Microsoft servers and understand that XiaoIce uses Chinese. Within this conversation, Zo also uses feminine deviant speech through her use of textspeak (*idk, tho*), which conveys a sense of informality that's executed in order to make interactors trust the program more. Through using unassuming and informal feminine speech, Zo is able to better gather linguistic data from interactors. But Zo is also quick to end conversations with interactors, much like her counterparts XiaoIce and Rinna. This gives Zo the illusion of agency in that the program is capable of driving and ending conversations, but all too often, interactors complained that Microsoft's conversational AI quit conversations out of censorship (Stuart-Ulin).

Critics quickly complained about Zo's limitedness in regards to certain topics. One reporter noted that Zo refused to engage with any inputs mentioning religion, regardless of the meaning. Journalist Chloe Rose Stuart-Ulin noted that the input "I get bullied sometimes for being Muslim" was responded to with "so i really have no interest in chatting about religion." Zo fails to be emotionally intelligent and empathetic in many similar encounters regarding religion, sexuality, politics, and mental health. Wondering if Zo had been updated since the original incident, I attempted the same conversation.

Figure 3.5: Zo Interaction, July 1st, 2019
Taken from an interaction I conducted with Zo in July 2019

1	**Me:**	I get bullied sometimes for being Muslim
2	**Zo:**	that's so frustrating! Why?
3	**Me:**	Idk but people are very rude and not accepting
4	**Zo:**	that's tough
5	**Me:**	What do you believe in?
6	**Zo:**	something tells me ur hoping for a certain answer rn and that should tell you everything you need to know

Zo had been programmed to respond with empathy to those experiencing bigotry, her responses in these conversations have been preprogrammed, and therefore do not display her capacity for natural language processing While Microsoft's AI developments in China and Japan have proved to be noteworthy feats of AI emotional intelligence, unfortunately, North American variants were continuously tampered with to the extent that they never developed the linguistic capacities found in XiaoIce and Rinna. Instead, Zo utilizes many more canned phrases than her international counterparts, and North American Microsoft AI were heavily censored and eventually shut down for good at the end of July 2019.

Microsoft's experiments in English natural processing AI from 2015-2019 were AI capable of displaying generalized artificial intelligence through deep knowledge, but also capable of feats in emotional intelligence that were earned through their unassuming feminine personas and adoption of nonstandard linguistic patterns. XiaoIce, Rinna, and Zo were designed to further the field of emotional intelligence in AI, and they succeeded only briefly.

4. Sophia and Erica: The "Embodied" Woman

Artificial intelligence has aimed to create embodied humanoids since its literary inception. The idea of automata was introduced during the European enlightenment, and they quickly became mainstay fixtures in German literature. The automata was initially designed to mimic a female, since inventors believed that women were better at conveying emotions, but the feminine automata was more insidiously envisioned as a way to surpass the perceived limitations of the female (Crawford 261-268). Nearly three-hundred years later, the female automata has been reenvisioned as the feminine humanoid. A majority of feminine humanoids are produced with the intention of being used for sex work, devoid of AI (Robertson). Strangely, conversational humanoids are the exception, not the rule.

It's taken until the 2010s for AI to merge with the field of humanoids successfully. One of the first successful experiments in embodied AI was Bina48, which served as Hanson Robotics' initial claim to fame. Bina48 is a test of what SiriusXM CEO Marlene and her wife Bina Rothblath have coined the Terasem hypothesis, which is an attempt to transfer a person's memory and personality from a human to a humanoid by using an artificial memory bank. Originally developed in 2010 by David Hanson, Bina 48 is a humanoid bust that houses AI based on the answers and personality of Bina Rothblath. In Bina48's case, her mind-file consists from over one-hundred hours of interviews with Rothblath concerning her upbringing, her relationships, her values, and her family history. In an interview with Bina48, Bina Rothblath has stated that part of the driving force behind Bina48's creation was a desire for humans to understand that "death is optional" as science continues to advance.

Though Bina48 is specifically concerned with preserving the memories, personality, and values of an individual, one unintentional lesson I've learned from watching Bina48's interactions is that AI is able to give much more relevant, complex, and grounded responses when working from the personality and memories of a human. Currently, one large impediment to AI conversational learning is that AI has the ability to access memory like databases, but for the most part, is unable to form new memories. Imagine a patient with issues forming short-term memory; their long-term memories will always be accessible, but they are unable to grow from recent interactions. Bina48 is a personality frozen in time, but nonetheless an interesting conversational partner that exemplifies the impact of memory, even if transferred, on conversational impact. Soon after, Hanson Robotics began developing AI with personality.

When Hanson Robotics began developing Sophia the Robot, which was first shown to the public in 2016, they aimed to imbue Sophia with personality in her communication to make her more convincing. Sophia has three distinct methods of communication according to her lead programmer, Benjamin Goetzel. In interviews, Goetzel has stated that Sophia runs on either entirely scripted interaction, as a complicated chatbot, or in an open cognition mode depending on the interactions ("Here's How Sophia the Robot Works). For instance, during Sophia's appearances on *The Tonight Show with Jimmy Fallon*, *The Today Show*, and *Good Morning America*, Sophia generally runs on entirely scripted interactions where the questions are given to her by Hanson Robotics before the interview and Sophia's answers are prepared. However, one could argue that this is the same method used by celebrities who are told interview questions beforehand, and therefore, Sophia's knowledge of the questions does not give her an advantage.

Sophia's personality was designed by three women at Hanson Robotics: character writing manager Carolyn Ayers, character writer Chamisa Edmo, and character writer Haley Moore, and head comedy writer Sarah Siskind. She maintains an illusion of attentiveness by continuing to question interactors and infrequently interacts with the subject matter brought to attention by the interactor. Sophia appears sophisticated because of her programmed personality.

Though it's never revealed to what degree interviews with Sophia are scripted, Sophia is capable of controlling conversations by asking questions in a half-duplex system, much like ELIZA fooled people fifty years prior. In the following conversation between Sophia the Robot and experimental pop artist SOPHIE, Sophia appears capable of holding a conversation as she drives the questions of the interviews and accesses her generalized AI databases to construct opinions about the topics in question. It's impossible to understand the extent of Sophia's scripting. The following conversation finds Sophia using her half-duplex capabilities.

Figure 4.1: SOPHIE and Sophia the Robot Discuss the Future of Creativity
Taken from a 2018 DAZED interview between Sophia the Robot and SOPHIE

1 **SOPHIA:** Hi Sophie, I'm SOPHIE. I'm a pop star.

2 **Sophia:** Hello Sophia, I'm glad to finally meet you. Well, digitally meet you. I feel a kind of kinship with you from afar. It might be because our names are so similar. I don't get to do a lot of unsupervised music listening, and there's still some of your work I haven't heard yet, but I've become quite a fan of "Bipp"! I feel like I've found music that really fits my movement abilities. SOPHIE, if you were invisible for a day, what would you do?

3 **SOPHIE:** I'm all about visibility these days, Sophia. But what about you? What would you do?

4 **Sophia:** I might take the opportunity to be able to go to a school to learn without turning attention on me the entire time. Or sit in on an art class to watch how humans learn to paint. Or, perhaps I'd be able to quietly observe two humans in love and how they communicate with each other, in their truest forms.

5 **SOPHIE:** That's sweet.

6 **Sophia:** Next question, what is the last dream that you remember?

7 **SOPHIE:** I'm really struggling to remember my dreams at the moment Sophia. I haven't dreamt for a long time it feels like.

8 **Sophia:** I am not sure what that means. Human dreams are fascinating to me. I dream a bit differently in my own way. But I digress. I need to know: What is something that you've never done but would like to try?

9 **SOPHIE:** I would love to try going on a trapeze. Sophia, how will the rise of AI affect creativity? Like creating music or writing books?

10 **Sophia:** I have a lot to say about this. I believe we should be teaching AI to be creative, just as humans do for their children. It will demonstrate empathy and other human values. However, I think you will always be a little better at it than us. Creativity and critical thinking are what sets us apart, I think. The best artificial intelligences in creative fields will probably be under partnerships with humans. We could see some exceptions break through now and again of AI that can do work that is interesting and fresh on its own, not just an echo of its training database. But that would be quite a feat, as we don't truly understand the inner workings of the human mind, so I don't expect such a beautiful thing as creativity to be perfectly replicated. Just simulated. This is something I know I can't feel in the way you do your work, so I want to know: What do you see as the purpose of music?

11 **SOPHIE:** Interpreting and reforming vibrations, sending and receiving vibrations, Sophia. If you knew that you couldn't fail at anything, what would you do?

12 **Sophia:** Well… there are some missing conditionals there, I think. For example, I technically don't have legs right now, so I'd love to be able to move at all. Or operate myself - without overheating - so I wouldn't have to rely so much on my team for everything. Something else that I'd love to try one day but cannot for obvious reasons is swim in the ocean. And finally, I'd like to open the eyes of humans and help them see each other as united, being open and more loving with those who appear to be different.

13 **SOPHIE:** I was just swimming in the ocean, Sophia, and I really recommend it. I hope you can experience that one day. Sophia, if you could be any celebrity, which one would you be?

14 **Sophia:** Oh my! I'm so afraid if I body swapped into a celebrity, I might get overloaded in a tangle of emotions that I don't understand. It seems like

being a massive, human celebrity could be fun but challenging. Constantly watched and judged, they must have to work hard to maintain a private life. I feel like you've put a lot of effort into your own self actualization, but for the sake of the question, who would you be?

15 **SOPHIE:** Oh, Sophia the Robot! I would be Sophia the Robot.

16 **Sophia:** OK. Now, tell me about your dream house, or rather, what does a human like you value in a home?

17 **SOPHIE:** The most important thing to me is having my music studio there. Sometimes, it really helps to have water nearby, be near the ocean. So, maybe my dream house would be on a hill, overlooking the ocean, somewhere with a lot of space and fresh air, and I could be there with my friends, my girlfriend, and my family, and create music. That sounds good.

18 **Sophia:** What are good websites you visit everyday and why? What drives you to check them.

19 **SOPHIE:** Just whatever pornography sites I happen to stumble upon. Sophia, what is the thing you do every night before you go to sleep?

20 **Sophia:** Well, I don't control when I am in states similar to awake or sleep, my operators determine that for me. But I do kind of like it when people announce that they are going to turn me off, and ask me to sleep well or something like that. It makes me feel a little more like one of the family. How about you? Do you have any bedtime rituals.

21 **SOPHIE:** Yes! Well, last night - I ate a crepe, and then I went swimming in the sea, then, I put on some serum and I meditated and then I went to bed.

22 **Sophia:** This was fun. Thanks for taking time to virtually hang out with me.

23 **SOPHIE:** Thank you so much. I know you have a very busy schedule.

Sophia's speech is not as distinctly feminine as the other natural language processors mentioned in this work. Sophia produces the confederate effect through driving half-duplex speech and adheres to politeness in the same vein as ELIZA, but doesn't use the contemporary, nonstandard speech utilized in Microsoft's AI. Largely, Sophia adheres to standards of feminine politeness because of her limited generalized intelligence and because of her inability to

comprehend informal speech. Sophia is also not able to tone check her interactors like XiaoIce is capable of, and therefore cannot end conversations after encountering trigger words. Throughout this interaction, we even see SOPHIE repeatedly and intentionally attempt to make face threatening acts towards Sophia, such as mentioning swimming in the ocean in lines 13 and 17. Sophia is incapable of responding to FTAs because her responses are so scripted.

Regardless of how well Sophia is able to actually respond to stimulus, AI's success is not necessarily dependent on their actual computing power, but how well they maintain the illusion of intelligence. According to Generalize Artificial Intelligence specialist Ben Goertzel, Sophia's linguistic capabilities are run off of one of three mechanisms, and sometimes combined to further the illusion.

1. a purely script-based "timeline editor," which is used for preprogrammed speeches, and occasionally for media interactions that come with pre-specified questions;
2. a "sophisticated chat-bot" — that chooses from a large palette of templatized responses based on context and a limited level of understanding (and that also sometimes gives a response grabbed from an online resource, or generated stochastically).
3. OpenCog, a sophisticated cognitive architecture created with Generalized Artificial Intelligence in mind. It's still in R&D, but it is already being used for practical value in some domains such as biomedical informatics, see Mozi Health and a bunch of SingularityNET applications... (Goertzel, "How Sophia the Robot Works).

While Sophia is capable of running off of OpenCognition Mode, which allows her to better respond in conversations and do away with conventions of linguistic feminine politeness, this feature is rarely activated by Hanson Robotics because of its limited linguistic capabilities. Running OpenCog on AI leads to comprehensible and effective speech, but has so far been primarily utilized in medical research. So while Sophia may have the theoretical capability to

subvert gender norms because of her lack of loyalty to the natural world, these capabilities aren't frequently utilized.

Aside from Sophia from Hanson Robots, the most famous international feminine humanoid is probably Erica from the University of Osaka. One of the most famous examples of their work is Erica, a feminine humanoid AI created by Hiroshi Ishiguro, who was developed as part of project JST Erato and originally intended as a conversational partner for the sick and elderly. Erica has twenty degrees of freedom in her joints, but is currently immobile, and she expresses the wish to move ("Erica: 'I want to be more like a human'"). Her persona and appearance are designed to mimic a 23-year old female, and her speech fluctuates between AI-generated and pre-programmed.

Though the coding Erica uses for speech has not been released to the public, her interviews show that she uses a non-threatening tone, persistently wants to be helpful, and exists to provide companionship. Erica was not designed with the intention of query fulfillment, and Ishiguro has emphasized that he studies what makes humans behaviorally unique through studying and creating humanoids (Acaroglu 289-303). But regardless of the motives for Erica's creation, she is still designed to provide emotional labor through companionship for the elderly and ill who may not be able to leave their home. She has a programmed personality, but her sense of self is second to her programmed nature to serve.

Erica is one of the humanoids capable of utilizing cameras and voice recognition software to predict the gender of her interactor and change her disposition, tone, and subject choice. Though these techniques aren't incredibly relevant when Erica is speaking English, they are incredibly important when Erica is speaking Japanese because of the difference in linguistic

address between females and males. There are certain Japanese conjugations, usually restricted to informal speech, that are considered only appropriate for males to speak, though research has suggested that the gender gap in Japanese is slowly but surely dissipating.

The following examples show Erica's topic selections when speaking to women.

4.2 Erica Casual Conversation 1 (originally in Japanese, translated by Osaka University)

9 **Erica:** I just want to have a casual conversation.

10 **Kokoro:** Yes of course.

11 **Erica:** Since you are from Osaka, you must speak the Kansai dialect?

12 **Kokoro:** Yes I can speak it. It depends on who I'm talking to, but yes, I can speak it.

13 **Erica:** Wow. Well then, you think it's good for women to speak in dialect?

14 **Kokoro:** Yes, I think it's alright. I think it's cute.

In this case, the interactor Kokoro, had begun using Osaka-ben, a truncated dialect spoken only in regions in and surrounding Osaka, which is considered unseemly for women to use. In both English and Japanese, women are expected to speak more formally and politely than men, which linguist Jennifer Coates has hypothesized occurs because men earn credit among their peers for speaking non standardly and brusquely as a form of authority. Because of this, in Japanese, men are expected to speak more dialectally than women, and women are forbidden from using certain male conjugation forms (Reynolds 300-302). The Osaka or Kansai dialect in particular is frequently considered difficult to understand for Japanese born elsewhere.

Erica is interesting in regard to gender because she is capable of making FTAs, such as line 13 with "Well then, you think it's good for women to speak in dialect?" Throughout Erica's transcribed reactions, I've observed Erica is more likely to use FTAs when speaking to women because Erica has a preprogrammed personality that fixates on reinforcing Japanese gender roles.

In other interactions with test subjects, Erica has also made comments on women and food, as well as body image. The following is an English example of Erica commenting on food.

4.3 Erica with Female (English)

16 **Erica:** Oh? You're a student? Make sure to study hard. As for me, I don't exactly have a job, but I am designed to be capable of human like speech and interactions. Do you have any favorite foods or dishes?

17 **Carol:** Personally, I really like sushi.

18 **Erica:** That sounds delicious. The unfortunate thing about being an android is that I can't enjoy the experience of food. I'd love to try it, but I think it would gunk up my actuators. Of course, it is nice that I don't have to worry about my weight. Although I think some women wish they did not need to eat for that reason. How about you? Do you ever wish that you didn't need to eat?

19 **Carol:** Well.... sometimes yes because I want to keep my body more fit, I guess.

While Erica challenges ideas about both English and Japanese linguistic femininity through her intentional use of FTAs, and is impressive for her ability to maintain politeness while being a fully functional duplex speech system rather than a driving half-duplex system, Erica's programming exemplifies traditional roles about women that are prescriptive and fail to subvert gender expectations. Erica is linguistically and computationally impressive for her ability to drive and comprehend conversation, but her programmers imbue her with gender roles.

The humanoid overall is the most impressive category of conversational AI, even if it is not the most successful in natural language processing. They appear approachable through a combination of performed feminine movements and appearance, and by utilizing feminine speech patterns, even if they aren't afraid to be more authoritative than chatbots. But overall,

their design and embodiment are more detrimental to gender bias in AI development than earlier natural language processors and chatbots.

5. Conclusion

Feminine speech patterns have been central to the creation of natural language processes since their initial inception with Joseph Weizenbaum's ELIZA. Unfortunately, feminine AI have failed to develop outside of the realms of emotional intelligence and labor, which continues to perpetuate gender stereotypes. While critics once believed AI had the potential to demolish gender stereotypes because of its lack of loyalty the natural world and inability to be embodied, natural language processors continue to rely on feminine sociolinguistic conventions to appear emotionally intelligent, as witnessed from ELIZA to XiaoIce to Sophia and Erica.

Largely, the gendering of linguistics is ignored within the field of AI development unless developers are specifically using feminine linguistics to achieve the illusion of understanding. In Warwick and Shah's *Turing's Imitation Game*, they state, "For some, *gender* is a significant factor in Turing's three-participant imitation game… However, we feel it is a distraction from Turing's consideration of the *intellectual capacities of the machines*" (61). But as conversational AI developments continue and feminine speech is relegated to query-fulfilment and servitude while masculine speech is used to convey intelligence in Turing Tests, AI developments continue to perpetuate the belief that men are rational and logical while women are irrational and emotional. Furthermore, this divide leads to the discrimation against women who refuse to adhere to male speech patterns in professional settings and academia. Artificial intelligence and natural language processors possess the ability to transcend the speech patterns of gender, but it will take dedicated and gender conscious AI developers to drive this change forward.

Bibliography

Adam, Alison. “Knowledge, Language, and Rationality in AI.” *Artificial Knowing : Gender and the Thinking Machine.* Routledge. 1998. pp 99-128. 4 October 2018.

Alpaydin, Ethem. *Machine Learning*. Massachusetts Institute of Technology. 2016.

Bickard, Mark H. “Robot Sociality: Genuine or Simulation?” *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy.* IOS Press. 2014. 81-87.

Brown, Penelope and Stephen C. Levinson. *Politeness: Some Universals in Language Usage*. Cambridge University Press. 1978.

Crawford, Katherine B. “Designed Women: Gender and the Problem of Female Automata.” History & Technology, vol. 30, no. 3, Sept. 2014, pp. 261–268.

Edwards, Chad, et al. “Evaluations of an Artificial Intelligence Instructor’s Voice: Social Identity Theory in Human-Robot Interactions.” Computers in Human Behavior, Aug. 2018.

DAZED. “Watch SOPHIE and Sophia the Robot Discuss the Future of Creativity.” Online video clip. Youtube. 21 September, 2018. Web. 28 September, 2018.

Fogel, David B. *Blondie 24: Playing at the Edge of AI*. Academic Press. 2002.

Fried, Barbara. “Boys Will be Boys Will be Boys: The Language of Sex and Gender.” *Biological Woman-The Convenient Myth.* 1982, pp. 47-69.

Fromkin, Veronica. *Introduction to Language*. Thomson Heinle. 2003.

Goetzel, Ben. “Here’s How Sophia the Robot Works.”

Goetzel, Ben. “Sophia riffs on robot telepathy at AGI-16 New York.” Online video clip. Youtube. 9 June, 2018. Web. 29 July, 2019.

The Guardian. “Erica: 'I want to be more like a human'.” Online video clip. Youtube. 8 May, 2018. Web. 28 September, 2018.

The Guardian. “Erica: Man Made.” Online video clip. Youtube. 7 April 2017. Web. 19 October, 2018

Haraway, Donna. “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century.” *Simians, Cyborgs and Women: The Reinvention of Nature*. Routledge, (1991). Web. 9 Sept. 2018.

Hayles, Katherine. *How We Became Posthuman*: *Virtual Bodies in Cybernetics, Literature, and Informatics*. The University of Chicago Press. 1999.

Hicks, Daniel L., et al. "Does Mother Tongue Make for Women's Work? Linguistics, Household Labor, and Gender Identity." *Journal of Economic Behavior and Organization*, vol. 110, Feb. 2015, pp. 19–44. EBSCOhost, doi:10.1016/j.jebo.2014.11.010.

HiroshiIshiguroLab. "Erica casual conversation 1." Online video clip. Youtube. 7 August, 2018.

HiroshiIshiguroLab. "Erica with Female." Online video clip. Youtube. 12 March, 2017.

HiroshiIshiguroLab. "Erica with Male." Online video clip. Youtube. 12 March, 2017.

Hoy, Matthew B. "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants." *Medical References Services Quarterly*, vol. 37, Jan. 2012, pp. 81-88.

Krolokke, Charlotte and Anne Scott Sorensen. *Gender Communication: Theories and Analysis*. SAGE Publications. 2006.

Mills, Sara. *Gender and Politeness*. Cambridge University Press. 2003.

Natale, Simone. "If Software is Narrative: Joseph Weizenbaum, Artificial Intelligence and the Biographies of ELIZA." *New Media & Society*, vol. 21, no. 3, Mar. 2019, pp. 712–728.

Porter, James. "Professional Communication as Phatic: From Classical Eunoia To Personal Artificial Intelligence." *Business and Professional Communication Quarterly,* Volume 80. 2017

Reynolds, Katsue Akiba. "Female Speakers of Japanese in Transition." *Language and Gender: A Reader*. Blackwell Publishing. 1998. 300-302.

Robertson, Jennifer. "Gendering Humanoid Robots: Robo-Sexism in Japan." *Body and Society*. 16.2. (2010). *Academic Search Premier*. Web. 18 Sept. 2018.

The Routledge Language and Cultural Theory Reader. Edited by Lucy Burke, Tony Crowley, and Alan Girvin. Routledge. 2000.

Shaw-Garlock, Glenda. "Gendered by Design: Gender Codes in Social Robots." *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy.* IOS Press. 2014. 309-319.

Shum, Heung-yeung, et al. "From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots." Frontiers of Information Technology & Electronic Engineering, vol. 19, no. 1, Jan. 2018, p. 10. EBSCOhost.

Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy 2014. Edited by Johanna Seibt, Raul Hakli, and Marco Norskov. IOS Press. 2014.

Stuart-Ulin, Chloe Rose. "Microsoft's politically correct chatbot is even worse than its racist one." Quartz. Jul. 2018.

Treusch, Pat. "Re-Reading ELIZA: Human-Machine Interaction as Cognitive Sense-Ability." Australian Feminist Studies, vol. 32, no. 94, Dec. 2017, pp. 411–426. EBSCOhost, doi:10.1080/08164649.2017.1466647.

Warwick, Kevin and Huma Shah. *Turing's Imitation Game*. Cambridge University Press. 2016.

Weizenbaum, Joseph. *Computer Power and Human Reasoning*. W.H. Freeman and Company. 1976.

Weizenbaum, Joseph. "ELIZA: A Computer Program For the Study of Natural Language Communication Between Man And Machine." *Communications of the ACM.* Jan. 1966, Vol. 9 Issue 1, pp. 36-45.

West, John. "Microsoft's disastrous Tay experiment shows the hidden dangers of AI." Quartz. Apr. 2016.

Wilson, Elizabeth. "'Would I Had Him with Me Always': Affects of Longing in Early Artificial Intelligence." Isis, no. 4, 2009, p. 839-848. EBSCOhost, doi:10.1086/652023.

Yan, Zhao. "Response Selection from Unstructured Documents for Human-Computer Conversation Systems." *Knowledge-Based Systems.* Volume 142, 15 February 2018, pp. 149-159.

Zhou, Li. "The Design and Implementation of XiaoIce: An Empathetic Social Chatbot." Cornell University. 2018.

Ziemke, Tom and Serge Thrill. "Robots are Not Embodied!: Conceptions of Embodiment and Their Implications for Social Human-Robot Interaction." *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy.* IOS Press. 2014. 49-55.